·数智医疗·

基于人工智能的临床决策支持系统早期临床评估的透明化报告

雷芳[1]，杜亮[1，2，3，4]，董敏[1]，刘雪梅[1*]

1.610041 四川省成都市，四川大学华西医院华西期刊社
2.610041 四川省成都市，四川大学华西医院华西人工智能医疗器械临床试验中心
3.610041 四川省成都市，四川大学华西医院医疗器械监管研究与评价中心
4.610041 四川省成都市，四川大学华西医院中国循证医学中心
*通信作者：刘雪梅，编审；E-mail：liuxuemei@wchscu.cn

【摘要】 随着人工智能（AI）在医学领域的广泛应用，越来越多基于AI的临床决策支持系统应用于临床诊断与筛查等。早期临床评价对于评估AI临床决策支持系统的临床性能、安全性和人为因素具有重要意义，并为进一步的大规模临床试验奠定基础，但相关临床研究报告的透明度和完整性尚待提高。2022年5月基于AI的临床决策支持系统早期临床评估报告指南（DECIDE-AI）正式在线发表。本文将结合DECIDE-AI中的AI相关报告项目及相关文献，探讨基于AI的临床决策支持系统早期临床评估的透明化报告，以期帮助我国AI临床决策支持系统的开发者和研究者更好地理解与应用相关指南，提高AI临床决策支持系统早期临床评估报告的透明度。

【关键词】 人工智能；决策支持系统，临床；早期临床评估；DECIDE-AI；透明化
【中图分类号】 R-05 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0668

Transparent Reporting of the Early Clinical Evaluation of Decision Support Systems Based on Artificial Intelligence

LEI Fang[1]，*DU Liang*[1，2，3，4]，*DONG Min*[1]，*LIU Xuemei*[1*]
1.West China Periodicals Press of West China Hospital，Sichuan University，Chengdu 610041，China
2.Clinical Trials Center for Artificial Intelligence Medical Device，West China Hospital，Sichuan University，Chengdu 610041，China
3.Medical Device Regulatory Research and Evaluation Center，West China Hospital，Sichuan University，Chengdu 610041，China
4.Chinese Evidence-Based Medicine Center，West China Hospital，Sichuan University，Chengdu 610041，China
**Corresponding author：LIU Xuemei，Senior editor；E-mail：liuxuemei@wchscu.cn*

【Abstract】 With the wide application of artificial intelligence（AI）in the medical field，more and more AI-based clinical decision support systems have been applied in the clinical diagnosis，screening，and other fields. Early clinical evaluation is important for evaluating the clinical performance，safety，and human factors of AI-based clinical decision support systems，and laying the foundation for large-scale trials. However，the transparency and integrity of the clinical reports need to be improved. The Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence（DECIDE-AI）was officially published online in May 2022. Based on this guideline and relative literature，this paper explores the transparent reporting of early-stage clinical evaluation of AI-based decision support systems，in order to help developers and researchers of AI-based clinical decision support system better understand and apply the relevant guidelines，and improve the reporting transparency of early clinical evaluation of AI-based clinical decision support systems.

基金项目：国家自然科学基金资助项目（72074161）
引用本文：雷芳，杜亮，董敏，等. 基于人工智能的临床决策支持系统早期临床评估的透明化报告［J］. 中国全科医学，2024.［Epub ahead of print］. DOI：10.12114/j.issn.1007-9572.2023.0668.［www.chinagp.net］
LEI F，DU L，DONG M，et al. Transparent reporting of the early-stage clinical evaluation of decision support systems based on artificial intelligence［J］. Chinese General Practice，2024.［Epub ahead of print］.



【Key words】 Artificial intelligence；Decision Support Systems，Clinical；Early clinical evaluation；DECIDE-AI；Transparency

近十年来，为提高医疗服务质量和效率，人工智能（artificial intelligence，AI）已广泛应用于临床诊断、筛查等方面[1-5]。尽管有研究指出，在临床前的计算机模拟研究中一些基于 AI 的算法准确性与临床专家相似[6]，但 AI 临床决策支持系统在临床决策中的应用受到复杂的人为和环境因素影响，从算法到临床应用需要经过多个阶段的评估，包括计算机模拟试验、临床早期评估和临床随机对照试验[7-8]。其中，AI 临床决策支持系统的早期临床评估的目的是，在大型临床随机对照试验前，于临床应用场景中对AI临床决策支持系统的临床性能、安全性和人为因素等进行验证，为以后的大规模临床试验奠定基础。因此，AI 临床决策支持系统的早期临床评估十分重要。

临床研究报告的透明化有助于提高试验的可靠性和结果的可重复性。此外，从伦理方面来讲，AI 决策的透明性和可解释性也是相当重要的[9]。目前，AI 相关的临床研究报告的透明度和完整性尚待提高。有系统评价研究指出，AI 相关的临床研究报告缺乏透明性，甚至有些研究本身存在较大缺陷，如未报告数据预处理信息、偏倚风险等，也有研究未评估模型适用性或未经外部验证[10]。AI 相关的临床研究报告规范可帮助研究者提高报告的完整性和透明度。2022 年 5 月基于 AI 的临床决策支持系统早期临床评估报告指南（Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence，DECIDE-AI）[11]正式在线发表，DECIDE-AI 主要针对早期临床评估，该研究阶段类似于药物开发的第 1/2 阶段，或外科创新技术 / 器械临床研究设计、开发、评价、长期随访（Idea，Development，Exploration，Assessment，and Long-term follow up，IDEAL）框架的第 2a/2b 阶段[12]。DECIDE-AI 共包含 27 个报告条目，其中 AI 相关报告条目 17 个（共 28 个子条目）、通用报告条目 10 个。DECIDE-AI 主要关注临床可用性、安全性和人为因素，因此该指南对试验评估环境、实施过程、人为因素等都提出了具体要求。

为推动我国 AI 临床决策支持系统早期临床研究及其报告的规范化和透明化，本文将结合 DECIDE-AI 中的 AI 相关报告项目及相关文献，从 AI 临床决策支持系统的预期用途、研究过程、伦理和数据共享几个方面探讨基于 AI 的临床决策支持系统早期临床评估的透明化报告，以促进报告规范的理解与应用，提高 AI 临床决策支持系统早期临床评估报告的透明度。

1 AI 临床决策支持系统的预期用途

基于 AI 临床决策支持系统的早期临床研究报告需要对系统的目标用户、临床应用场景和预期效果进行适当描述。AI 临床决策支持系统融合了多种跨学科知识和技术，其对用户有一定的要求，包括专业、培训水平、计算机技术等；此外，对于 AI 临床决策支持系统的临床应用场景和预期效果的描述，包括其适用范围、可能给现有临床路径带来的改变和影响等，不仅有利于增加临床医生或患者对 AI 临床决策支持系统的了解，提高其对 AI 临床决策支持系统的信任和满意度，同时可促进 AI 临床决策支持系统的传播和实施。

2 研究过程的透明化报告

尽管行业协会或组织均发布政策，要求在卫生保健和研究领域开发 AI 项目时保持透明性，但由于其固有的复杂性，AI 仍缺乏透明性，包括算法缺乏可解释性、所有阶段数据不够透明、数据偏见以及模型版本化不够透明等。AI 的透明性对各利益攸关方都意义重大，包括研发者、使用者、审批和监管机构等。下面主要从参与者、AI 临床决策支持系统、实施、安全和故障、人为因素 5 个方面对研究过程透明化报告进行阐述。

2.1 参与者

AI 诊断工具相关的 Meta 分析显示，大部分研究存在较高的患者选择偏倚风险[13-14]。患者选择偏倚不仅影响 AI 诊断模型的推广，更重要的是，可能会产生系统错误，即输出错误的结果。因此，应加强患者招募过程的透明化报告，从患者和数据层面报告纳入与排除标准，以及患者招募数量的确定方法。AI 临床决策支持系统的参与者除患者以外，还包括用户，用户与 AI 临床决策支持系统之间的交互也将对系统性能和研究结果产生影响，因此不仅需要报告用户的招募过程以及纳入与排除标准，还需要描述用户熟悉 AI 临床决策支持系统使用所需时间和工作量，但 DECIDE-AI 指南不推荐从数据层面排除用户。此外，研究者也可在临床试验注册平台进行预注册，这是提高临床研究过程透明化的重要途径，同时这也符合《赫尔辛基宣言》、WHO 和国际医学杂志编辑委员会对临床试验注册的要求。

2.2 AI 临床决策支持系统

AI临床决策支持系统可能会经历多次迭代和更新。有研究表明，与决策支持过程相匹配的算法可在一定程度上减少系统的固有偏倚[15]。除了算法，AI 临床决策支持系统的另一大核心便是数据，大多数 AI 临床决策

支持系统都需要大型数据集来进行训练和验证，因此数据的获取、输入以及处理都在一定程度上决定了 AI 临床决策支持系统的性能[15-16]。在 AI 临床决策支持系统输出方面，输出内容和呈现方式一定程度上影响了用户对系统的使用和接受度[17]，而良好的人机交互建立在合理的用户界面设计基础上。为此，作者在研究设计和撰写研究报告时，都应仔细考虑并描述算法类型和版本、输入数据相关信息以及输出内容和交互界面设计等。

2.3 实施

在实施方面，与 AI 临床试验报告统一标准（Consolidated Standards of Reporting Trials - Artificial Intelligence，CONSORT-AI）和 AI 临床试验方案报告指南（Standard Protocol Items：Recommendations for Interventional Trials - Artificial Intelligence，SPIRIT-AI）不同，DECIDE-AI 除了要求报告 AI 临床决策支持系统的输出如何应用到临床决策，还要求清楚描述评估环境，如医疗中心的类型和规模、软硬件支持等。这些环境因素可能会影响数据的获取或测量，进一步影响输出结果。研究者还需要报告最终的临床决策是如何形成的、用户使用系统情况及对现有诊疗和护理流程的重要改变，这可能在一定程度上反映了 AI 临床决策支持系统的可用性和临床实用性。

2.4 安全和故障

安全问题是医学 AI 的首要问题，它涉及各种法律和伦理问题[18]，并贯穿整个临床研究和实践过程。因此，在大型临床研究前，对 AI 临床决策支持系统进行安全评估尤为重要，以避免造成灾难性后果。当然，在使用 AI 临床决策支持系统的过程中，安全风险或故障不仅涉及系统本身，软硬件支持和用户使用情况也可能会产生错误。如何识别这些故障、如何评估患者安全风险、如何最小化患者伤害，都需要研究者提前制定好相应的安全风险预案，并详细报告。

2.5 人为因素

人为因素的评估主要围绕用户的使用情况，即可用性。可用性是指一个产品可以被指定用户使用，并在指定使用情景中，有效地、有效率地、满意地达到指定目标的程度[19]，这直接影响了用户对系统的接受度。所以，研究者可采用量表、问卷调查表等方式，从满意度、完成任务的效率和工作量等方面进行评估。如广泛用于系统可用性研究的系统可用性量表（System Usability Scale，SUS）[20]可快速评估可用性，该量表包含 10 个条目，每个条目的得分为 0~4 分。另外，也可采用净推荐值（net promoter score，NPS）[21]，即用一个简单的问题来评估用户的满意度。事实上，生活中类似的 NPS 调查也很常见，如“你向朋友推荐 XX 产品的可能性有多大？”或“你会向你的朋友推荐 XX 产品吗？”。关于如何选择恰当的方法，可以参考相关指导文件[22]。

DECIDE-AI 指出，报告中需要对结果是否支持 AI 临床决策支持系统的预期用途、安全提示进行充分阐述，这充分体现了早期临床研究的目标，即临床应用性和安全性。在临床应用性方面，研究者应根据研究结果客观描述系统的临床表现，避免夸大其应用潜能，这将对后期是否继续优化和完善 AI 临床决策支持系统以及是否进一步进行临床验证提供重要的证据支持，避免不必要的人力、物力和财力的浪费。对系统安全性的讨论可促进 AI 临床决策支持系统优化，并增强用户或患者对系统的信任。

3 伦理

AI 临床决策支持系统在促进医疗服务质量提升和生产力提高的同时，也带来了一系列伦理挑战，比如患者的数据隐私、系统的准确性、算法偏见等。尽管在医疗 AI 中，医疗保健数据集通常在开发人员使用前被去识别化，软件也有可能通过与其他数据之间的交叉引用来重新识别底层个体。2022 年 3 月发布的我国首个国家层面的科技伦理治理指导性文件《关于加强科技伦理治理的意见》提出，科技伦理是开展科学研究、技术开发等科技活动需要遵循的价值理念和行为规范，是促进科技事业健康发展的重要保障[23]。2023 年 2 月由国家卫生健康委、教育部、科技部和国家中医药局联合发布的《涉及人的生命科学和医学研究伦理审查办法》强调了伦理审查的重要性，并指出“涉及人的生命科学和医学研究应当尊重研究参与者，遵循有益、不伤害、公正的原则，保护隐私权及个人信息”[24]。在 AI 临床决策支持系统的开发和应用中需要始终坚持“伦理先于技术”的原则，必要时可采用适当的技术方法或伦理学方法以降低相关伦理风险，如隐私计算技术可使数据“可用而不可见”，避免隐私数据泄露。AI 相关的伦理内容与传统医学伦理并不完全一致，因此应当加强 AI 相关医学伦理的宣传与审查。学术期刊作为研究结果的重要传播平台，应发挥好宣传与审查作用。然而，根据我们的调查，目前很少有国内外医学期刊更新投稿指南或作者须知，补充 AI 相关的医学伦理要求[25]。

4 数据共享

如前所述，算法和数据是 AI 临床决策支持系统的两大核心，分享相关数据一方面可供读者验证研究结果，另一方面也可促进交流、完善算法代码，同时这也符合透明化和可重复性原则。数据获取是验证和共享 AI 产生的科学知识的方式。数据共享不仅是为了促进科学发展，同时也是为了提高科学的严谨性。由于 AI 研究与代码和数据密不可分，因此应该更加强调代码和数据共

享，深化 AI 临床试验的价值，预防错误、欺诈或偏倚。因此，除了声明利益冲突外，作者还应该声明相关数据和代码是否可获取以及如何获取。随着开放科学时代的到来，数据共享将变得越来越重要。

5 总结

随着 AI 在医学领域的广泛应用，为增加研究报告的完整性和透明性，相应的报告规范（指南）不断地被开发出来。DECIDE-AI 为 AI 临床决策支持系统的研究报告提供了一个框架，以帮助开发人员和研究者报告开发、实施和评估过程中的关键信息。该指南主要关注了早期临床评估的 3 个方面：小规模临床使用的临床实用性、安全性和人为因素。为进一步促进 DECIDE-AI 的传播，加强相关研究报告的规范性、完整性和透明性，同时为作者在研究设计、方案制定和研究注册过程中提供参考标准，建议期刊将 DECIDE-AI 纳入稿约或作者须知。但由于 AI 临床决策支持系统在临床应用中可能面临多种多样的问题和挑战，因此研究人员在运用该指南时，需要根据实际情况进行调整和改进。同时，由于制定该指南的专家大多来自英国，所以可能会有一定的局限性。尽管如此，该指南给 AI 临床决策支持系统的相关研究及报告提供了一定的方向，有利于促进 AI 在医疗领域的可持续发展。

作者贡献：雷芳负责论文构思与设计、初稿撰写；杜亮负责论文修订、质量控制及审校；董敏负责资料收集与整理、论文修改；刘雪梅负责论文设计与修改，对文章整体负责。

本文无利益冲突。

雷芳：https://orcid.org/0000-0003-1660-8937

刘雪梅：https://orcid.org/0000-0003-3201-5568

参考文献：

[1] KERMANY D S，GOLDBAUM M，CAI W J，et al. Identifying medical diagnoses and treatable diseases by image-based deep learning [J]. Cell，2018，172（5）：1122-1131.e9. DOI：10.1016/j.cell.2018.02.010.

[2] TANG A，TAM R，CADRIN-CHÊNEVERT A，et al. Canadian association of radiologists white paper on artificial intelligence in radiology [J]. Can Assoc Radiol J，2018，69（2）：120-135. DOI：10.1016/j.carj.2018.02.002.

[3] Pelcyger Ben. Artificial intelligence in healthcare：Babylon Health & IBM Watson take the lead [EB/OL]. [2023-09-19]. https://prescouter.com/2017/12/artificial-intelligence-healthcare/.

[4] SLOMKA P J，DEY D，SITEK A，et al. Cardiac imaging：working towards fully-automated machine analysis & interpretation [J]. Expert Rev Med Devices，2017，14（3）：197-212. DOI：10.1080/17434440.2017.1300057.

[5] DE FAUW J，LEDSAM J R，ROMERA-PAREDES B，et al. Clinically applicable deep learning for diagnosis and referral in retinal disease [J]. Nat Med，2018，24（9）：1342-1350. DOI：10.1038/s41591-018-0107-6.

[6] HANNUN A Y，RAJPURKAR P，HAGHPANAHI M，et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network [J]. Nat Med，2019，25（1）：65-69. DOI：10.1038/s41591-018-0268-3.

[7] VASEY B，NOVAK A，ATHER S，et al. DECIDE-AI：a new reporting guideline and its relevance to artificial intelligence studies in radiology [J]. Clin Radiol，2023，78（2）：130-136. DOI：10.1016/j.crad.2022.09.131.

[8] PARK Y，JACKSON G P，FOREMAN M A，et al. Evaluating artificial intelligence in medicine：phases of clinical research [J]. JAMIA Open，2020，3（3）：326-331. DOI：10.1093/jamiaopen/ooaa033.

[9] PLOUG T，SUNDBY A，MOESLUND T B，et al. Population preferences for performance and explainability of artificial intelligence in health care：choice-based conjoint survey [J]. J Med Internet Res，2021，23（12）：e26611. DOI：10.2196/26611.

[10] TORNERO-COSTA R，MARTINEZ-MILLANA A，AZZOPARDI-MUSCAT N，et al. Methodological and quality flaws in the use of artificial intelligence in mental health research：systematic review [J]. JMIR Ment Health，2023，10：e42045. DOI：10.2196/42045.

[11] VASEY B，NAGENDRAN M，CAMPBELL B，et al. Publisher Correction：reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence：decide-AI [J]. Nat Med，2022，28（10）：2218. DOI：10.1038/s41591-022-01951-8.

[12] HIRST A，PHILIPPOU Y，BLAZEBY J，et al. No surgical innovation without evaluation：evolution and further development of the IDEAL framework and recommendations [J]. Ann Surg，2019，269（2）：211-220. DOI：10.1097/SLA.0000000000002794.

[13] PARKASH O，SIDDIQUI A T S，JIWANI U，et al. Diagnostic accuracy of artificial intelligence for detecting gastrointestinal luminal pathologies：a systematic review and meta-analysis [J]. Front Med，2022，9：1018937. DOI：10.3389/fmed.2022.1018937.

[14] LI X M，GAO X Y，TSE G，et al. Electrocardiogram-based artificial intelligence for the diagnosis of heart failure：a systematic review and meta-analysis [J]. J Geriatr Cardiol，2022，19（12）：970-980. DOI：10.11909/j.issn.1671-5411.2022.12.002.

[15] GURUPUR V，WAN T T H. Inherent bias in artificial intelligence-based decision support systems for healthcare [J]. Medicina，2020，56（3）：141. DOI：10.3390/medicina56030141.

[16] FAGHANI S，KHOSRAVI B，ZHANG K，et al. Mitigating bias in radiology machine learning：3. performance metrics [J]. Radiol Artif Intell，2022，4（5）：e220061. DOI：10.1148/ryai.220061.

[17] FUJIMORI R，LIU K，SOENO S，et al. Acceptance，barriers，and facilitators to implementing artificial intelligence-based decision support systems in emergency departments：quantitative and qualitative evaluation [J]. JMIR Form Res，2022，6（6）：e36501. DOI：10.2196/36501.

[18] O'SULLIVAN S, NEVEJANS N, ALLEN C, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery [J]. Int J Med Robot, 2019, 15(1): e1968. DOI: 10.1002/rcs.1968.

[19] International Organization for Standardization. ISO 9241-11: 2018 ergonomics of human-system interaction—Part 11: usability: definitions and concepts [EB/OL]. [2023-09-19]. https://www.iso.org/standard/63500.html.

[20] Brooke, J. SUS—a quick and dirty usability scale [EB/OL]. [2023-09-19]. https://hell.meiert.org/core/pdf/sus.pdf.

[21] REICHHELD F F. The one number You need to grow [J]. Harv Bus Rev, 2003, 81(12): 46-54, 124.

[22] British Standards Institution. BS EN 62366-1: 2015+A1: 2020 medical devices. Application of usability engineering to medical devices [EB/OL]. [2023-09-19]. https://www.bsigroup.com/en-GB/standards/bs-en-62366-12015a12020/.

[23] 中共中央办公厅，国务院办公厅.关于加强科技伦理治理的意见[A/OL].(2022-03-20)[2023-09-19]. https://www.gov.cn/zhengce/2022-03/20/content_5680105.htm.

[24] 国家卫生健康委，教育部，科技部，等.涉及人的生命科学和医学研究伦理审查办法[A/OL].(2023-02-27)[2023-09-19]. http://www.nhc.gov.cn/qjjys/s7946/202302/c3374c180dc5489d85f95df5b46afaf5.shtml.

[25] 雷芳，杜亮，董敏，等.关于人工智能背景下医学期刊应对医学伦理问题的思考[J].编辑学报，2023，35(3)：263-267. DOI：10.16811/j.cnki.1001-4314.2023.03.005.

（收稿日期：2023-10-09；修回日期：2023-10-13）

（本文编辑：王世越）